MyWeather: The Social Media of Weather

An Honors Thesis (HONR 499)

by

Mitchel W. Hill

Thesis Advisor

Dr. David Call

Ball State University
Muncie, Indiana

December 2013

Expected Date of Graduation

December 2014

Abstract:

Weather is a conversation staple for humanity. It is always changing, giving people a fresh perspective to discuss, and it has remained as a hazard with severe weather outbreaks and unexplained processes. Founded on this idea, the subject of this creative Honors thesis is to establish a connection between social media and meteorology events, specifically creating a data source that could be spatially representative of current weather conditions. Atmospheric processes can vary greatly between locations and can be difficult to monitor from satellite and radar only. With 65 million social media users in the United States alone, this could be a valuable resource for tracking the weather and monitoring real time changes.

Acknowledgements:

Special thanks are extended to Dr. David Call for advising through the weather aspects of this project as well as numerous other avenues relating to the major.

I would also like to thank Nathan Foss, my partner on the project, whose time and effort on computer coding made this project possible.

By creating a program to synthesize atmospheric data with social media posts, a connection can be established between these two media. This merging of technologies can provide helpful insight into atmospheric changes, as well as what people find notable in the weather. The creation of a weather social media platform has a number of steps that are a part of the creative process. This thesis outlines the more difficult step in this process, data collection. There are two requirements we placed on the data we were to collect with this application, the data needed a spatial component and needed to be accurate. Spatially linked data is necessary for cartographic display, which was the goal for this data collection. It is important that this spatial data be as accurate as possible to eliminate any misconceptions and false information. These criteria raised concerns with verification methods, cross-reference data availability, spatial extent functionality, and selection of necessary variables.

The first obstacle I was faced with in this project was finding a method of verifying the posts that would run through our social media site. This source would need to be reliable, providing accurate and specific data with a subject range that could follow likely weather-related social media posts. The initial subject data we were checking for included temperature, precipitation, and wind variables. The data would also need to update, operating within small time parameters to account for rapid changes in atmospheric activity. Ideally these updates would be continuous, but considering data storage capabilities and download times, a more realistic measure was hourly updates as a maximum. Finally, this data would need to have spatial information stored jointly with the data variables so a function could recognize the verification station closest to a social media post.

A logical source for the necessary variables is aerodrome routine meteorological reports (METAR). METAR is the international accepted code form for the transmission of surface observed weather data. This code is endorsed and monitored by the World Meteorological Organization (WMO) and the International Civil Aviation Organization (ICAO). It can be created manually, but the emergence of automated weather stations has rendered this form obsolete. The automated weather stations that transmit this data in the United States are operated and maintained jointly by the Federal Aviation Administration (FAA) and the National Oceanic and Atmospheric Association (NOAA). Needless to say, the source of this data can be considered credible. It accomplishes the other requirements as well, transmitting anywhere from 11 to 54 different weather-related variables every hour.

The next step in the verification source selection process was identifying the type of weather station that was the most reliable and transmitted the proper variables in a manner that was usable in computer coding. There are three main types of automated weather stations in the United States, which is the area this project is restricted to. These stations include the Automated Weather Observing System (AWOS), the Automated Surface Observing System (ASOS), and the Automated Weather Sensor System (AWSS). AWOS and AWSS stations are exclusively ran by the FAA; however ASOS stations are jointly operated by the FAA and the NOAA. The ASOS stations on average provide a wider range of atmospheric measured data than the other two station types including the expanded variables we settled on for social media post recognition, making this station type the best fit for our project. AWSS stations have similar variables as the ASOS stations and were the secondary choice. AWOS stations come in multiple varieties depending on the variables the station is equipped to monitor. The only variation of this station type employed in our site is the AWOS III P/T meaning it is the most recent version of the station with precipitation and thunderstorm detection equipment. These decisions were based upon the weather variables we wanted to represent in our program. Those variables included high temperature, low temperature, rain, snow, three hour pressure tendency, sky condition, visibility, wind speed, funnel cloud, and tornado. Combinations of these variables were used to further define the conditions, for which they could occur, such as snow requiring temperatures below 2 degrees Celsius, overcast cloud conditions, or falling three hour pressure tendency if snow is not reported on the last METAR.

With the variation and types of weather observation data decided upon, the next problem to consider was where this data would come from. This question is two-fold, dealing both with how the data would get from the observation sites and where the observation sites are. The first part was answered by my partner when he found METARs that updated automatically in XML format, a coding language he could use in an interpretative process. The second decision that needed to be made was what U.S. observation locations would be utilized. A middle ground would need to be established between a large enough spatial representation for accurate verification while managing the time it would take to connect a post to the database. A method was developed to solve for this issue to effectively reduce the number of stations necessary for accuracy within our program.

The social media posts to our weather site would likely be from areas with internet connections, whether wired or cell tower provided. These services are generally provided more reliably in centers of higher population density. The first check for automated weather stations was for those near population centers greater than 2500 persons per square mile. Automated weather stations take direct measurements of a 10 mile radius for visibility, thunderstorm monitors, and lightening discriminators. Surface observations can be extended for a station and inferred for a region for roughly another 10 miles. A twenty mile zone was used as reference for further weather station decisions, so if an area had more than 40 miles in any direction without a weather station then a weather station in the region was added. This was then cross checked for highly visited areas with outside activities that did not have a representative year round population. The majority of these areas are National Parks and recreation areas, such as Glacier National Park. These criteria condensed the 28,000 automated weather stations across the contiguous United States including Alaska and Hawaii to a working database of 506 of mostly ASOS stations and limited AWSS or AWOS III P/T stations where necessary. This reduced number of stations accomplishes the goals of the program while not overloading the system with extraneous data.

Once the station locations were selected, the parameters provided by each model needed to be chosen and paired with possible social media posts. Everyday weather is not normally worth mentioning, so the parameters tested in the verification function would be related to severe weather and extreme changes. These atmospheric processes would encompass temperature fluctuations, precipitation, wind, fog, thunderstorms, and tornadoes. Common words and phrases used to describe these changes were the written into the program code to match the social media posts with a METAR. The variables defined for cross-checking were then used jointly with keywords from a post. Once this connection was made, or the post was determined to be weather-related by the program, it was checked against the nearest METAR to see if it agreed.

Temperature fluctuations were easy to identify using METARs; however, parameters needed to be set for more descriptive terms like "hot" or "cold." These parameters were set based on what an average person would call warm or cold using normal centralized air settings. Wind parameters were handled in a similar fashion, where days with high winds were set at a certain value to verify social media posts. Fog is a straightforward notation within a METAR. The other two variables have more complex notation. Precipitation can fall in a variety of forms, and a

METAR distinguishes between them all with different symbolization. The major forms identified in the program included rain, snow, hail, sleet, and mist. The more technical term symbolizations like "graupel" are not verified in the program. Storm related indices posed more of a challenge for verification. While thunderstorms and lightning have their own weather station apparatus for detection, this data is posted in raw text form within the remarks section of a METAR, which was not specially formatted in the XML language. Separate code had to be written to translate the remarks section, which also included tornado information. Tornadoes form rapidly and are at the micro scale, making them a difficult atmospheric process to track in a spatial environment. Tornado verification is essential to early warning systems, and MyWeather posts could provide another avenue, accompanied with radar data, to warn for tornadoes. Therefore, this information could be crucial; however, it is not always available in a timely fashion.

Weather is always in motion and developing. Often times the hourly updates received through the METAR reports do not reflect current conditions as accurately as the social media posts do, particularly towards the end of a cited period. To account for this with some of the more time sensitive parameters, certain indicators were written into the code to accompany the actual notation. Changes in pressure can indicate imminent weather changes, for instance a drop in pressure over a measured time interval can indicate possible storm activity. Cloud cover is another precondition we used to verify precipitation and storm activity, specifically the overcast notation used in METAR. Another indication of changes is funnel cloud readings in the remarks section, which would be used to verify any tornado posts within the site. Tornadoes are difficult to distinguish on radar. Generally, rotation can be seen, but this does not mean that a tornado has touched down. Tornado touchdown must be verified visually. MyWeather could be helpful in aiding to this task.

Once my partner finished the coding for the proposed application, the site went live for Ball State campus. This gave us the opportunity to work out any bugs we encountered with login procedures or post updating. When these things had been addressed, we requested server space through the Ball State Computer Science department to test the site's national functionality and speed. These tests proved to be positive, with a post coming from Florida on a mobile device that did not present an abnormal amount of lag. The posts we received came from the 60 accounts registered with the site providing 70 different weather updates. Of the 70 posts we received in our two week testing period, 23 verified as being true weather statements. Table 1 illustrates the verified weather posts, demonstrating a sampling of the types of posts we could expect to see with a permanent application.

A few bugs were discovered in the program during this testing period. One was not having a way for testing against climate differences. This was prevalent during Thanksgiving when temperatures were in the upper 40s and posts were received about the abnormal warmth. This did not verify for our parameters, ideally this would be corrected with changing climate

Latitude	Longitude	Content	Keyword	Time
40.4095	-85.5736	It is balls cold.	cold	11/24/2013 0:02:15
41.878	-88.0163	According to my phone, it\'s super cold	cold	11/25/2013 18:12:43
42.1695	-87.9588	Snowing here!	snow	11/25/2013 18:21:45
40.4398	-86.0893	Snow flurries this morning!	snow	11/27/2013 20:38:41
28.0222	-81.7329	It is 66 degrees, windy yet sunny here	wind	11/27/2013 20:48:32
40.4661	-88.9034	Bloomington, IL - 9:38 AM - 11/28/2013 -	wind	11/28/2013 15:42:32
39.2959	-85.9508	COLD.	cold	11/30/2013 1:15:38
39.8453	-84.1122	I am visiting family in Huber Heights,	snow	11/30/2013 2:52:31
37.9568	-87.3515	Really cold, but dry. Lots of wind.	cold	11/30/2013 4:33:41
42.7325	-84.5555	It is particularly cold outside and very	cold	11/30/2013 7:09:24
42.6735	-82.9165	There\'s a layer of clouds covering the	cold	11/30/2013 15:46:43
41.0996	-85.0635	The weather is pleasant in Fort Wayne,	cold	11/30/2013 22:12:34
38	-97	I\'m in West Milton, OH, and it\'s cold.	cold	12/1/2013 5:39:46
40.2369	-85.4679	Over break I was in La Porte, IN and the	snow	12/2/2013 3:31:05
40.2023	-85.4082	The weather\'s been alright, it\'s cold but	cold	12/2/2013 3:55:32
40.2519	-84.5151	It\'s supposedly 48 in lovely Oregonia,	cold	12/2/2013 19:53:58
40.1768	-83.0792	It\'s 44 and really foggy here in	fog	12/3/2013 16:49:55
40.2023	-85.4082	It was in the 30s this morning when I	cold	12/3/2013 17:39:23
40.2023	-85.4082	Temp is moderately cold. Wearing long	cold	12/3/2013 17:50:38
40.4725	-85.3422	It\'s fairly overcast today. It\'s not as	cold	12/3/2013 18:10:17
42.9742	-85.9311	Grand Rapids, MI - It\'s 33 degrees and	rain	12/4/2013 3:54:03
38.2813	-85.8401	it\'s 51 in Bloomington, IN! Also, still	fog	12/4/2013 14:03:23

Table 1: Verified MyWeather social media posts

parameters for each day. A similar issue was faced when posts only referred to the temperatures in degrees rather than a qualitative explanation. Additional code would need to be added to account for these values as well. The final difficulty experienced with these posts was updates that came from outside the spatial scope of our project. One post was received from the United Kingdom, which did not verify due to the location being outside an acceptable range from an automated weather station.

These findings were very promising for possible expansion of this undertaking. Many obstacles were faced with the synthesis of data and comparison of unpredictable weather events with unpredictable social media posts. The verification processes we implemented to maintain accuracy also present the employment of exciting application methods. An example of this would be our original plan for the system, which was to display the information spatially on an interactive map where you could view pertinent weather information in your region. These could include assisting in advanced warning systems or providing human environment perspective to atmospheric anomalies. Whether for technical purposes of personal enjoyment, MyWeather has the potential to provide quick and accurate information that is not always available through conventional methods.

Resources:

U.S. Department of Transportation: Federal Aviation Administration. *Surface Weather Observation Stations*. Retrieved from: http://www.faa.gov/air_traffic/weather/asos/

U.S. Census Bureau. *2010 Census Data.* Retrieved from: http://www.census.gov/2010census/data/